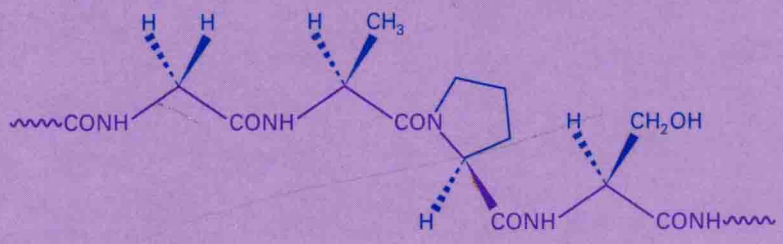
OXFORD CHEMISTRY PRIMERS

Amino Acid and Peptide Synthesis

John Jones



OXFORD SCIENCE PUBLICATIONS

7

Amino Acid and Peptide Synthesis

John Jones

The Dyson Perrins Laboratory and Balliol College, University of Oxford

OXFORD NEW YORK TOKYO
OXFORD UNIVERSITY PRESS
1992

Oxford University Press, Walton Street, Oxford OX2 6DP

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Berlin Ibadan

Oxford is a trademark of Oxford University Press

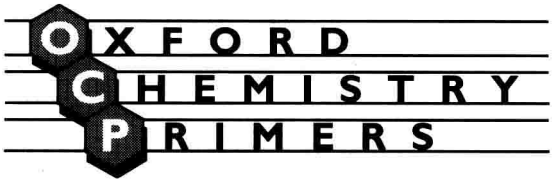
Published in the United Sates
by Oxford University Press, New York

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Jones, John, 1942–
Amino acid and peptide synthesis / John Jones.
(Oxford Chemistry Primers; 7)
Includes bibliographical references and index.
1. Amino acids — Synthesis. 2. Peptides — Synthesis. I. Title.
II. Series.
QD431. J623 1992 547.7'50459 — dc20 91–44596

ISBN 0–19–855669–1
ISBN 0–19–855668–3 (Pbk)

Printed in Great Britain by
Information Press Ltd, Oxford



SERIES EDITOR
STEPHEN G. DAVIES
The Dyson Perrins Laboratory, University of Oxford

Oxford Chemistry Primers

1	S. E. Thomas	*Organic synthesis: the roles of boron and silicon*
2	D. T. Davies	*Aromatic heterocyclic chemistry*
3	P. R. Jenkins	*Organometallic reagents in synthesis*
4	M. Sainsbury	*Aromatic chemistry*
5	L. M. Harwood	*Polar rearrangements*
6	I. E. Markó	*Oxidations*
7	J. H. Jones	*Amino acid and peptide synthesis*

Series Editor's Foreword

The control and regulatory mechanisms of biological processes are dependent on peptides and proteins which are derived from α-amino acids. There are now many medicines in use that are peptides or derivatives which are used, for example, as antibiotics, or to control blood pressure, or as anti-cancer agents. For this reason α-amino acid and peptide chemistry is central to organic chemistry and their synthesis is a core topic of most courses.

Oxford Chemistry Primers have been designed to provide concise introductions relevant to all students of chemistry, and contain only the essential material that would normally be covered in an 8–10 lecture course. In this seventh primer of the series, John Jones provides an excellent, easy to read account of amino acid and peptide synthesis aimed at second and final year students. This primer will be of interest to apprentice and master chemist alike.

Stephen G. Davies
The Dyson Perrins Laboratory, University of Oxford

Acknowledgement

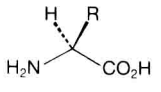
Although the coverage of amino acid synthesis is new, the greater part of this book is a distillate of the chapters on fundamentals which appear in my recent work *The chemical synthesis of peptides* (OUP, 1991). I am grateful for permission to use this material again.

J. H. J.
Oxford
July 1991

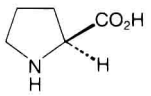
Contents

1 Introduction

Proteins are natural polymers which are assembled under nucleic acid control from a menu comprising nineteen L-α-amino acids of general structure **1** and the 'imino' acid L-proline (**2**): see Table 1.1. Amide or



(1)



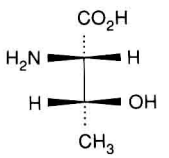
(2)

Table 1.1
The proteinogenic amino acids (**1**)

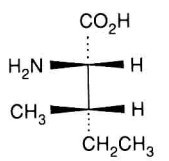
Amino acid*	–R
Alanine	$-CH_3$
Arginine	$-(CH_2)_3NHC(=NH)NH_2$
Asparagine	$-CH_2CONH_2$
Aspartic acid	$-CH_2CO_2H$
Cysteine	$-CH_2SH$
Glutamine	$-(CH_2)_2CONH_2$
Glutamic acid	$-(CH_2)_2CO_2H$
Glycine	–H
Histidine	$-CH_2$(4-imidazolyl)
Isoleucine**	$-CH(CH_3)CH_2CH_3$
Leucine	$-CH_2CH(CH_3)_2$
Lysine	$-(CH_2)_4NH_2$
Methionine	$-(CH_2)_2SCH_3$
Phenylalanine	$-CH_2Ph$
Serine	$-CH_2OH$
Threonine***	$-CH(CH_3)OH$
Tryptophan	$-CH_2$(3-indolyl)
Tyrosine	$-CH_2$(4-hydroxyphenyl)
Valine	$-CH(CH_3)_2$

*All the proteinogenic amino acids belong to the same, L-, stereochemical series, and all save one are designated (2*S*); the way the Cahn–Ingold–Prelog sequence rules work makes L-[but (2*R*)-]-cysteine an apparent anomaly. This minor confusion has reinforced innate conservatism, and so far ensured the survival of the L-D terminology in the field.
**(2*S*, 3*S*)
***(2*S*, 3*R*)

Some specialized proteins contain residues which are not in the Table. Several involved in blood clotting and calcium metabolism, for example, contain γ-carboxyglutamic acid residues, with R=$-CH_2CH(CO_2H)_2$, which bind Ca^{2+} tightly. Such proteins are still nevertheless biosynthesized exclusively from those listed here, with additional side-chain functionalization being introduced after backbone assembly.



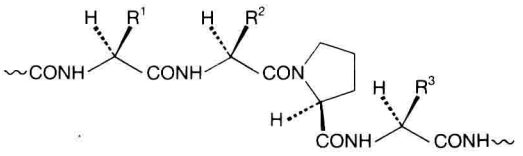
L-[2*S*,3*R*]-threonine



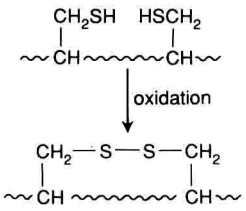
L-[2*S*,3*S*]-isoleucine

The non-proteinogenic diastereoisomers with the opposite configurations at C-3 are called L-*allo*-threonine and L-*allo*-isoleucine respectively.

'peptide' bonds link the building blocks, giving macromolecules (**3**) with

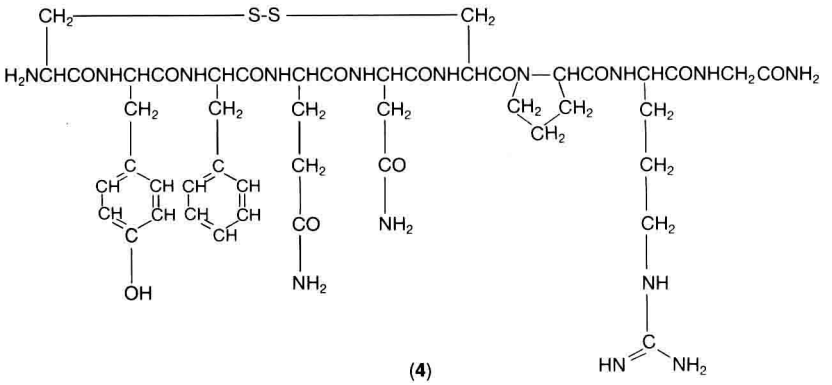


(**3**)



Disulphide bridge formation

'polypeptide' backbones and side-chains containing a variety of simple functionalities according to the amino acids selected. These structures may be further elaborated by inter- or intra-chain covalent connection (most commonly by the formation of disulphide bridges between thiol side-chains, as in **4**); or by non-covalent association; or by the coordina-



(**4**)

tion of metal ions; or by the attachment of auxiliary components such as haem, carbohydrates, phosphate groups, or lipids; or by chemical modification in other ways, which include acetylation, hydroxylation, methylation, and carboxylation. Proteins may be acidic, basic, or neutral. They may be globular and soluble, or fibrous and insoluble. Their molecular weights span a range which extends from a few thousand to several million. They are present in abundance and diversity in every part of every living thing on Earth. Their indispensability was recognized long before anything about them could be understood in present-day terms, and is the origin of the name 'protein', which is derived from the Greek $\pi\rho o\tau\epsilon\iota o\sigma$, ranked first.

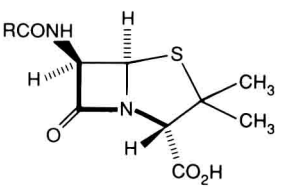
Peptides are molecules constructed in essentially the same way as proteins, but on a smaller scale. The terms 'peptide' and 'protein' are used very inconsistently in the literature of the subject. 'Protein' is appropriate in current usage for any macromolecule incorporating more than about fifty of the usual α-amino acid residues assembled under nucleic-acid direction, or for a close synthetic analogue. The classification 'peptide' is suitable for molecules with fewer amino acid residues than this, whatever the amino acids and however produced, even if there is extensive modification after assembly of the residues by peptide bond formation. A prefix is often applied to indicate the number of residues—e.g. tripeptide, three residues.

The full structural formula of vasopressin (**4**), a pituitary nonapeptide which was one of the first of the peptide hormones to be synthesized (1954), and which is, from today's vantage point, a very simple example, is too complex to be taken in at a glance. A system of highly abbreviated formulae has been developed to present such structures. In this system, each amino acid has a three-letter code, derived from its trivial name. A one-letter code has also been agreed for the proteinogenic amino acids (i.e. those found in proteins), but it is more useful in connection with amino acid sequence data than for peptide synthesis. The simplification which can be achieved with the three-letter code is well illustrated by comparison of the full (**4**) and abbreviated (**5**) formulae for vasopressin. Together with cryptic symbols for substituents, the three-letter codes can be used to formulate any amino acid or peptide derivative, including synthetic intermediates. An International Commission has recommended detailed conventions which are, with minor aberrations, universally observed. Their system, which will be used throughout this book, is outlined in Appendix A, along with lists of amino acid codes and substituent symbols. Fluency with the system will be assumed, and readers are advised to study Appendix A before going further. The abbreviations used for common reagents and solvents are also defined there.

Vasopressin (**5**)

Catalytic proteins—enzymes—mediate practically all the operations in the molecular business of biology. Biochemical balance is also regulated by hormones, the majority of which are proteins or peptides. Membranes contain proteins which control permeability, and help to pump solutes through against thermodynamic gradients. Many neuroactive peptides and proteins with complex interrelationships have been identified in the brain. Selectively toxic peptides and proteins are deployed by many species, either for defence against predators, or in aggressive competition for limited nutrient resources. The venoms of snakes, bees, and wasps, for example, are evil cocktails of this sort; the lethal bacterial toxin responsible for botulism is a protein of high molecular weight; the principal poisonous constituents of the notorious toadstool *Amanita phalloides* are complex cyclic peptides; and antibiotics such as penicillin are, though modified almost beyond recognition as such, nevertheless peptides of a kind. The toxic peptides of bacteria and fungi are of special chemical interest because they are structurally very varied, incorporating D- as well as L-configurations, and amino acids which are often quite different from those of proteins. Carrier proteins transport smaller molecules from one location to another. Haemoglobin, loading and unloading oxygen as required, is a familiar example. Skin, bone, hair, horn, tendon, muscle, feather, tooth, and nail are all largely proteinaceous, as are spiders' webs, larval cocoons, and antarctic fish antifreeze agents. The immune system, with which animals defend themselves against invasive parasites, employs proteins to recognize and reject foreign molecules, exercising exquisitely precise discrimination between self and not-self. Even genetics requires proteins, because although the genetic information is encoded in nucleic acid structures, the DNA is highly organized in compact association with basic proteins called

Actually there are many penicillins, of the general structure shown below. The bicyclic system is biosynthetically derived from the dipeptide unit whose backbone is emboldened.



That the 3D (tertiary) structure of a protein is determined by the amino acid sequence (primary structure) of the protein or a precursor is well proven (see Section 9.2.2.2), but it is not yet possible to predict tertiary structures from sequence data.

histones, and an array of other proteins is necessary for genetic expression and control. Furthermore, the genes dictate only protein amino acid sequences. These determine three-dimensional structure, from which all else flows. All the data necessary to specify that an organism will have the form and characteristics of an orchid, an oryx, or an orang-utan is written in the amino acid sequences of its proteins.

It follows that there are few aspects of biology which cannot be illuminated by experiments using peptides or proteins. Nor is the interest merely academic, for peptides and peptide-related structures can influence endocrine, neurological, immune, and enzymic processes with high specificity and prodigious potency. They therefore have varied applications or potential in medicine: in the regulation of fertility; the control of pain; the stimulation of growth; the therapy of cancer, cardiovascular problems, connective tissue diseases, digestive disorders, mental illness, and infections by pathogens. Several peptide drugs are in widespread clinical use, necessitating their manufacture by the kilo, and at least one company is said to derive an income of over two hundred million Swiss francs a year from its peptides. Such drugs all ultimately depend on the selective activation or inhibition of an existing part of the biochemical apparatus, by the specific recognition of, and interaction with, natural receptors or active sites—a universal theme of peptide and protein biochemistry. Then there is the prospect of synthetic peptide vaccines. If the immunologically significant parts of a protein can be mapped, then in principle they can be mimicked in a synthetic peptide vaccine, which can be used to immunize against the agent they belong to. This strategy, avoiding as it does the hazards and limitations of making and using vaccines in the traditional way, has already been shown to be feasible, and seems certain to become very important. And now it also begins to be possible to conceive a fundamentally new kind of approach to peptide and protein applications. Much remains to be discovered about the relationship between amino acid sequence, three-dimensional structure, and function, but an unrestrained imagination can look forward to the time when the chemist will be able to design and construct entirely unnatural peptides and proteins with predetermined novel properties.

Examples of synthetic peptide drugs include Synacthen^R (an adrenocorticotrophin fragment, for the therapy of rheumatic diseases, bronchial asthma, allergic disorders, skin conditions, etc); DDAVP^R (a vasopressin analogue, for the treatment of *diabetes insipidus*); Sandimmun^R (cyclosporin A, a cyclic peptide immunosuppressive drug used in organ transplantation) and Calsynar^R (salmon calcitonin, for the assessment and management of hypercalcaemia, Paget's disease, bone cancer, etc).

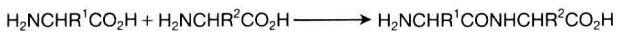
The demand for peptides and proteins is thus enormous, and rising all the time. Natural sources can provide a great variety, sometimes freely and in quantity, but the range is limited to what happens to be there, and more often than not only miniscule amounts can be isolated. The manipulation of biological systems by genetic engineering is an increasingly sophisticated art, and will no doubt become the principal means of manufacture for many natural and unnatural amino acid sequences. For example, insulin, a mini-protein of two chains and about fifty amino acid residues, which is essential for the treatment of *diabetes mellitus*, was formerly obtained from beef or pig pancreatic tissue. Because insulins from these sources are not quite identical with human insulin, some patients become sensitized against what are, to their systems, foreign substances. The obvious solution is to administer the human hormone. This demand could never adequately be met from natural sources, but some of the human insulin needed is now being prepared commercially

by recombinant DNA technology. Several other important peptides are being produced in the same way. However, each such instance requires developmental work, and the approach is at present impotent in the face of structural novelty. Chemical synthesis, on the other hand, is in principle applicable to any target. Since medicinal chemists commonly turn to unnatural analogues in search of metabolic stability, and the screening of large numbers of variants is generally necessary in the development of a new drug, chemical peptide synthesis is likely to remain an essential tool for the forseeable future. It can be estimated that at least five thousand people are engaged full-time in such work worldwide at the present time.

The aim of this book is to outline the principal chemical methods which are available for the synthesis of α-amino acids and peptides. Ruthless selectivity will be necessary, and there will be no space at all for technical matters (a deficiency greatly regretted, for this is a field in which the advances of the last three decades have been dependent on technology as well as chemistry). Similarly, even a rudimentary coverage of principles will leave little room for real examples. The subject and the molecules it has tackled with success are just too big. Furthermore, we cannot in the space available stray far from the proteinogenic amino acids and 'ordinary' peptides, although the synthesis of peptides containing other amino acids and non-protein structural features is of very great importance.

Nature synthesizes proteins by stepwise assembly from amino acid building blocks, and this is the way the chemist generally proceeds in the laboratory.

The formation of a dipeptide **6** having no side-chain functionality by the condensation of two amino acids with the elimination of water can be



(**6**)

Scheme 1.1. Conditions: dehydration.

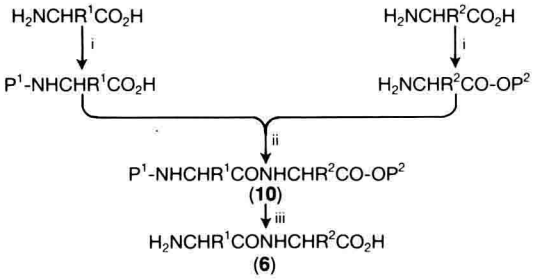
represented formally as in Scheme 1.1. It is clear that even if the chemistry of this conversion could be achieved, which in the naive manner shown it cannot, it would be ambiguous and uncontrolled. There being no means of distinguishing between the two amino groups or the two carboxy groups, the product **6** would be accompanied by the dipeptides **7**–**9**, together with the polycondensation products from all four possible

(**7**) (**8**)

(**9**)

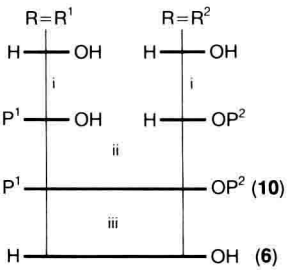
dipeptides. The controlled synthesis of **6** would entail the blockade (or 'protection') of the amino group belonging to one component (the 'carboxy component') and the carboxy group of the other (the 'amino com-

ponent'), followed by condensation (or 'coupling'), and finally removal of the blocking groups from the intermediate product **10** ('deprotection')—a minimum of four separate reactions (Scheme 1.2, otherwise represented



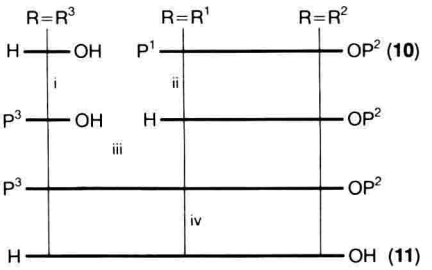
Scheme 1.2. Conditions: i, protection; ii, coupling; iii, double deprotection.

as in Scheme 1.3). For the conversion of **10** to a tripeptide **11**, there would be the additional requirement that P^1 should be removable without disturbing P^2, i.e. P^1 and P^2 should be 'orthogonal'. This would mean a minimum of eight separate reactions for the synthesis of **11**



Scheme 1.3. Conditions: i, protection; ii, coupling; iii, double deprotection.

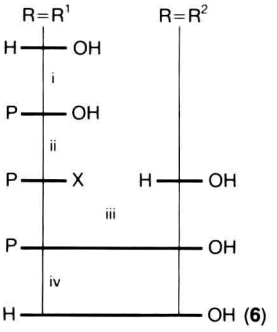
via **10** (Scheme 1.4) from free amino acids. Then there is the complication that the functional side-chains of the proteinogenic amino acids



Scheme 1.4. Conditions: i, protection; ii, selective deprotection; iii, coupling; iv, double deprotection.

all interfere with the reactions available for peptide bond formation to some degree, and usually require protection orthogonal to at least one of the α-function protection methods chosen, which limits choice and adds to the number of operations. An alternative approach is to separate the carboxy activation and coupling stages. In this case, the amino

component may be attached by a 'salt coupling', without the need to protect its carboxy group, provided exchange of activation between the free and activated carboxy groups does not intervene (Scheme 1.5). This



Scheme 1.5. Conditions: i, protection; ii, activation; iii, coupling, with the amino component in the salt form; iv, deprotection.

stratagem reduces the number of protecting group manipulations, but introduces an extra operation at the coupling stage, and may be disadvantageous for other reasons such as solubility.

Furthermore, all the transformations in a peptide synthesis must needs be performed without loss of stereochemical integrity at any of the chiral centres, if a product free of very similar, and possibly inseparable, diastereoisomeric contaminants is to be obtained. Lastly, the final removal of all the protecting groups must be carried out without destruction of the peptide backbone or other side-reactions. All of which makes the unambigious synthesis of large peptides sound unattainable. In practice things are not quite as bad as this. To start with, although unusual amino acids may have to be synthesized (Chapter 2) and appropriately derivatized (Chapters 3–5) before peptide synthesis can begin, the requisite derivatives of all the standard amino acids, protected and ready for incorporation, are commercially available. And the methodology of the subject has now been finely tuned, automated synthetic techniques have been developed, and modern purification technology can in fact cope with very complex mixtures.

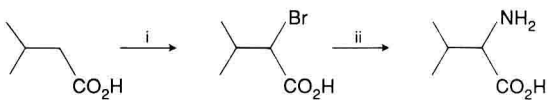
It is convenient to discuss the chemistry and merits of the main kinds of protection and of the reactions available for peptide bond formation in turn (Chapters 3–6), but it must be understood that no rigid rules can be laid down prescribing the best weapons or plan of attack for any particular synthesis. The protecting groups and coupling procedures must be able to work together without conflict, and the overall strategy and orchestration (Chapters 7–9) require judicious planning just as much as the individual steps call for care in the selection of the optimal methods.

2 α-Amino acid synthesis

The proteinogenic α-amino acids are produced industrially by fermentation methods and by chemical synthesis on a vast scale, reckoned for some in hundreds of thousands of tons per annum. Their principal application is as food additives, but they are incidentally cheap starting materials for laboratory work. The synthesis of α-amino acids at the bench is nevertheless an active field because of the demand for specifically labelled, unnatural and unusual amino acids. The need is almost always for a homochiral product, so assembly of the target without regard to α-chirality must be followed by resolution; alternatively an asymmetric synthesis must be employed; or an enantiospecific conversion of a freely available homochiral compound to the required α-amino acid must be achieved. Examples of all three approaches will be given.

2.1 General synthetic methods

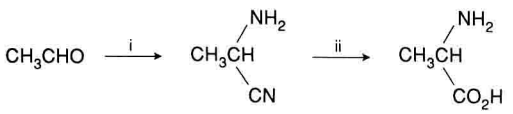
Many of the general methods for the synthesis of α-amino acids, including displacement reactions on α-halo acids (e.g. Scheme 2.1), the Strecker



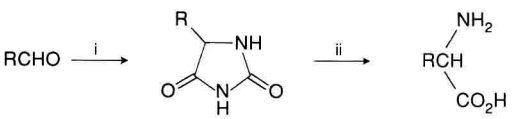
Stage i is the Hell–Volhard–Zelinsky reaction; it probably proceeds via the enol of the acyl halide, and thus only α-bromination takes place.

Scheme 2.1. Conditions: i, Br_2/PCl_3; ii, NH_3.

synthesis (e.g. Scheme 2.2), approaches through hydantoins (e.g. Scheme 2.3), and via oxazolones (e.g. Scheme 2.4), were developed in the early days of amino acid chemistry, but retain their importance. Although ammonia works well enough in conversions of very simple α-halo acids to

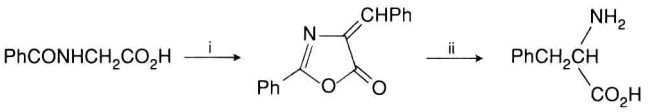


Scheme 2.2. Conditions: i, $NaCN/NH_4Cl$; ii, H_3O^+.



Problem: suggest a mechanism for stage i.

Scheme 2.3. Conditions: i, $KCN/(NH_4)_2CO_3$ (there are several alternatives for this ring-formation); ii, H_3O^+ or OH^-.

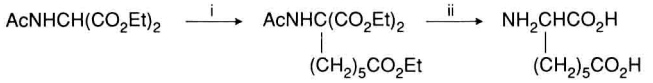


Scheme 2.4. Conditions: i, $PhCHO/Ac_2O/NaOAc$; ii, aq. HI/P/heat.

Stage i here is the Erlenmeyer azlactone synthesis (azlactone = oxazolone); it involves cyclodehydration of benzoylglycine to give an oxazolone, which has enhanced acidity at the CH_2 group and condenses easily with benzaldehyde.

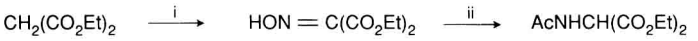
α-amino acids, potassium phthalimide (followed by strong acid hydrolysis of the intermediate phthalimido derivative: the Gabriel synthesis) or azide ion (followed by reduction) are superior reagents.

However, none of the above procedures is as frequently employed as that involving acylaminomalonates (e.g. Scheme 2.5). Although amino-



Scheme 2.5. Conditions: i, NaOEt, then $Br(CH_2)_5CO_2Et$; ii H_3O^+.

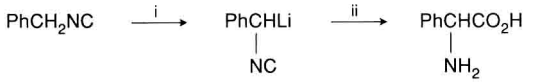
malonic acid and its α-alkyl derivatives are isolable, they are unstable with respect to decarboxylation, so that α-amino acids are produced directly under the vigorous conditions of the last stage. Diethyl acetamidomalonate is easily obtained as in Scheme 2.6, and a range of other



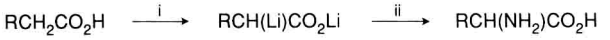
Scheme 2.6. Conditions: i, $NaNO_2/AcOH$; ii, $H_2/Pd(C)$, then Ac_2O.

acylaminomalonates can be made and applied analogously (see Scheme 2.10).

Two further general strategies are illustrated in Schemes 2.7 and 2.8.



Scheme 2.7. Conditions: i, Bu^nLi; ii, CO_2 at $-80\,^\circ C$, then H_3O^+.



Scheme 2.8. Conditions: i, Pr^i_2NLi; ii, NH_2OMe.

2.2 Resolution

The reactions illustrated in Section 2.1 all inevitably lead to racemic products. The traditional general approach to the resolution of racemates of all kinds is to derivatize with an optically active reagent, sep-